STATISTICAL METHODS USED TO

ESTABLISH BACKGROUND DATASETS USING SAMPLED DATA COLLECTED FROM DTLS, AND SURFACE AND SUBSURFACE SOILS OF THREE RBRAS OF THE TWO FORMATIONS

AND

COMPUTE ESTIMATES OF BACKGROUND THRESHOLD VALUES BASED UPON ESTABLISHED BACKGROUND DATASETS (WITH AND WITHOUT NONDETECT OBSERVATIONS)

FOR THE SANTA SUSANA FIELD LABORATORY INVESTIGATION

Prepared by

Anita Singh

October 2011

This page was intentionally left blank.

1.0 INTRODUCTION

This appendix provides a brief description of the statistical and graphical methods that were used to: 1) establish background datasets using the surface and subsurface soils data collected from three radiological background reference areas (RBRA) from the Santa Susana (Bridle Path RBRA) and Chatsworth (Rocky Peak and Lang Ranch RBRAs) formations; and 2) compute estimates of the background level concentrations/background threshold values (BTV) based upon the established background datasets thus obtained. This appendix will also describe the process and methods used to establish defensible background datasets free of outliers (Section 2.0). Sectoion 3.0 describes the methods used (in this Background Study Report) to compute BTV estimates for datasets with nondetects (ND) and without NDs. Section 4.0 discusses an alternative approach to compute statistics of interest including BTV estimates based upon radionuclide datasets consisting of ND observations. The alternative approach uses ND values as detected observations.

Statistical software packages ProUCL 4.1 (U. S. Environmental Protection Agency [EPA], 2011) and Scout 1.1 have been used to perform statistical analyses as summarized in the various radionuclide evaluations (Appendix A) of this Draft Background Study Report for the Santa Susana Field Laboratory (SSFL).

In the present study, univariate BTVs have been computed for more than 50 radionuclides. For each radionuclide, many onsite observations will be compared with its respective BTV estimate. On-site observations exceeding BTVs may require re-sampling (including step-out sampling) and/or cleanup activities. Since on-site observations versus BTV comparisons will be performed for multiple contaminants, the use of multivariate statistical methods is better suited to provide a balance between false positive and false negative error rates (Singh and Nocerino, 1995). However, theory and methods associated with multivariate and robust statistical methods are quite complex; and therefore are beyond the scope of the current SSFL Background Evaluation Report. Using the multivariate methods on multiple contaminants perhaps divided into sub-groups of highly correlated radionuclides (e.g., all uranium, all thorium), one will be computing multivariate BTV estimates represented by multivariate simultaneous ellipsoids. An onsite vector observation (e.g., for thorium) lying within the ellipsoid can be considered as coming from the same background population; and vector observations lying outside the BTV ellipsoid potentially represent impacted locations requiring further investigations. A much lesser number of comparisons will be made when using multivariate methods; additionally multivariate methods take correlations among the various analytes into consideration. These multivariate methods are available in the EPA software package, Scout 2008 (EPA 2009). Statistical details of the robust univariate and multivariate methods are summarized in the Scout 2008 User Guide (EPA, 2009) and in Singh and Nocerino (1995).

To keep the document simple and understandable by all interested readers, the Project Team and the stakeholders decided to use well-known and well- adopted univariate methods as described in the existing environmental literature (e.g., EPA Guidance Documents, 1992, 2000, 2002, 2006, 2009). The advantage of using univariate methods is that these methods are described in most of the environmental text books (e.g., Gilbert 1987) and various guidance documents; and environmental scientists are familiar with these univariate methods.

As noted in the appendices summarizing statistical analyses and BTV computations for the various radionuclides, all background datasets collected from the three RBRAs are fairly consistent with low variability; and not many outliers were identified in those background datasets. Keeping these observations in mind, the Project Team recommends the use of the Bonferroni inequality based upper simultaneous limits (Section 3.0) as estimates of BTVs for the various radionuclides included in this Background Study Report for the SSFL.

2.0 STATISTICAL METHODS USED TO ESTABLISH BACKGROUND DATASETS BASED UPON THE SURFACE AND SUBSURFACE SOILS DATA COLLECTED FROM THE THREE RBRAS OF THE TWO FORMATIONS

For each radionuclide, the objective is to establish a defensible background data set represented by a "single" population free of outliers potentially representing impacted observations. There is some concern among the stakeholders that some of the chosen RBRA locations might have been impacted by the site activities; therefore all identified outliers will be excluded from the background data set. When data from the various strata could not be merged, separate BTV estimates were computed for each stratum.

This section describes statistical methods that have been used to establish background datasets based upon the data collected from the surface and subsurface soils of the three RBRAs. The established background datasets thus obtained have been used to compute estimates of the BTVs for the various radionuclides considered in this SSFL Background Study Report. Several steps were followed to assure that the datasets collected from the three RBRAs were representative of the site-specific background free of impacted locations (e.g., with elevated activities). A brief discussion of the step-by-step procedure used is described below.

In order to compute reliable and defensible estimates of BTVs, it is necessary that the background dataset represents a single statistical population; and the dataset does not consist of observations potentially representing the impacted locations (outliers). A well-established background dataset represents a single statistical population consisting of observations collected from unimpacted locations. For background datasets representing mixture populations, separate BTVs are computed for each individual population (e.g., RBRA). Observations representing potential outliers are not included in the computation of BTV estimates. It should be noted that various upper limits used to estimate BTVs are meant to provide coverage for population observations and not for any of the parameters (e.g., mean) of the sampled population. It is not appropriate to use upper confidence limit (UCL) of the mean as estimates of BTVs.

Whenever possible, statistical results are supplemented with formalized graphical displays. Graphical displays provide added insight (e.g., presence of outliers, data distributions and patterns, mixture populations, visual comparison of two or more groups) into datasets that is not possible to visualize and understand by reviewing the estimates and test statistics such as Dixon and Rosner outlier test statistics, WMW and ANOVA test statistics, and upper limits used to estimate BTVs.

Proper Identification of Outliers

Dixon and Rosner outlier tests require that the data set follows a normal distribution. However, it should be pointed out that the presence of moderate to extreme outliers (e.g., lying farther away from the tails of the normal distribution) destroys the normality of the data set. Therefore, one may not use Dixon and Rosner test to identify moderate to extreme outliers (lying outside the tails of a normal distribution) which are inevitable in environmental applications. Dixon (1953) and Rosner (1975) tests were developed when computing power that we have today was not available. These outlier tests are not meant to process complex environmental data sets as considered in this SSFL Radiological Background Report. Moreover, due to masking, these classical tests often fail to identify all outliers present in a data set. The Dixon test can be used on data sets of size ≤ 25 and the Rosner test can be used on data sets of size ≥ 25. The Rosner test requires the user to know the number of outliers that may be present in a data set. If an incorrect number of outliers are specified, the Rosner test may not be able to correctly identify all outliers present in a data set.

The use of modern computer intensive robust methods and graphical displays is recommended to properly identify all potential outliers present in environmental data sets. The use of Q-Q plots to assess data distributions and to identify outliers is quite common (Gnanadesikan, 1977; Hoaglin, Mosteller, and Tukey, 1983; Singh and Nocerino, 1995; Johnson and Wichern, 2002) in the statistical literature. As with all other tests used in this Report, Dixon and Rosner test results are also supplemented by graphical displays (e.g., Box plots and formal Q-Q plots). Graphical methods do not suffer from masking effects. Scout software equipped with robust outlier identification methods was used to verify the proper identification of outliers. However, due to the complexity of robust methods, results obtained using those robust methods are not included in this Report.

An example illustrating the issues associated with Dixon and Rosner tests is considered next. Advantages of using graphical and robust outlier identification methods are also discussed using the same data set.

Example. Consider a normally distributed data of size 20. The formal Q-Q plot (with Shapiro-Wilk test statistic) assessing the normality of the data set is shown in Figure 1 below. Next an outlier was added to this normally distributed data of size 20. Figure 2 has the normal Q-Q plot based upon the data set with the outlier. It is noted that the presence of even a single outlier causes the data set to be non-normally distributed. However, the Dixon test when used on the data set of size 21 identified the single outlier present in the data set. As mentioned earlier, the Dixon test suffers from masking and often fails to identify multiple outliers present in a data set. Next, a couple of more outliers are included in the data set. The Q-Q plot based upon 23 data points is shown in Figure 3. It is obvious from Figure 3 that there are 3 values which are well-separated from the rest of the data set and can be considered outlying with respect to the main dominant population represented by the original data set of size 20.

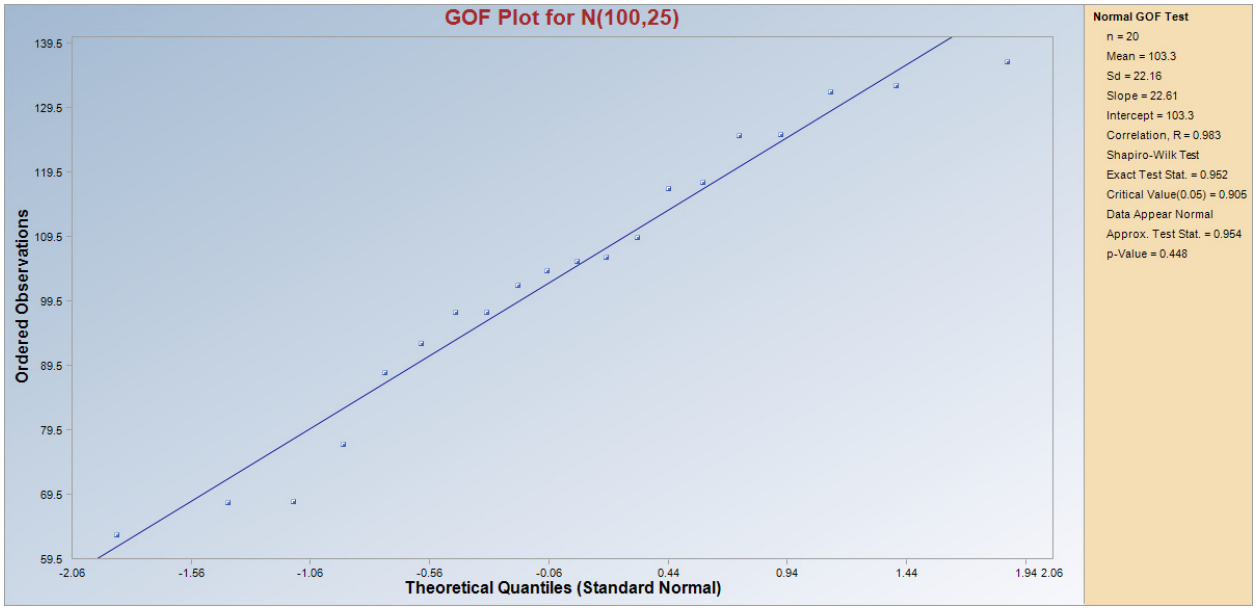

Figure 1. Formal Q-Q plot with Shapiro-Wilk Test Results - Assessing Normality of the Data Set.

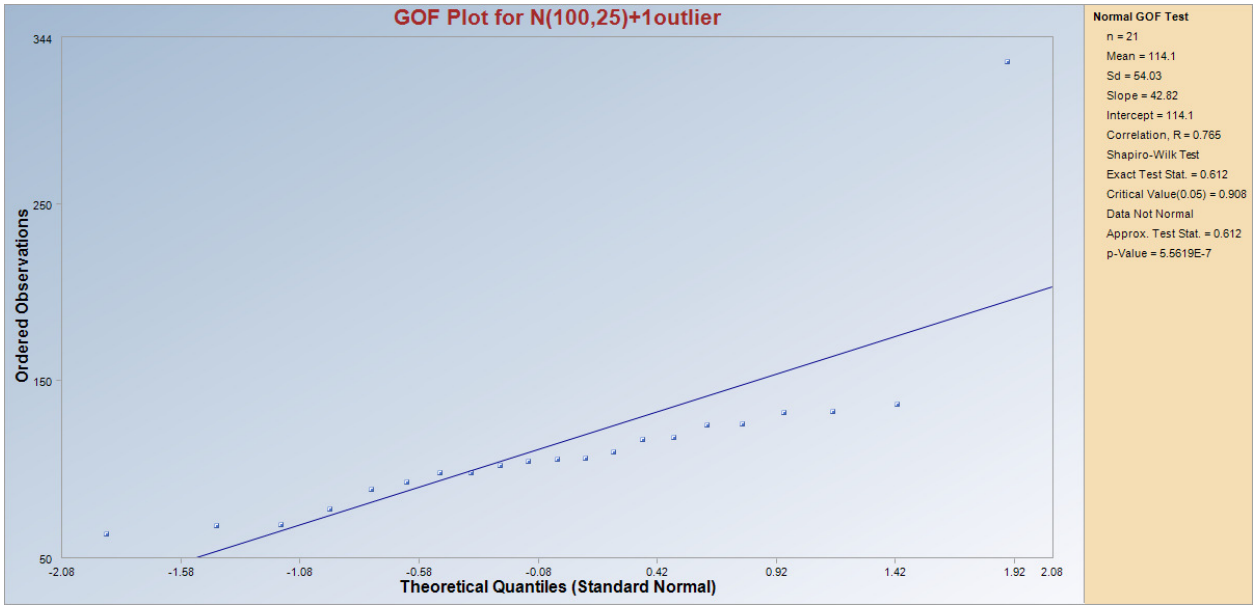

Figure 2. Formal Q-Q plot with Shapiro-Wilk Test Statistic and Indentifying Outliers Present in the Data Set.

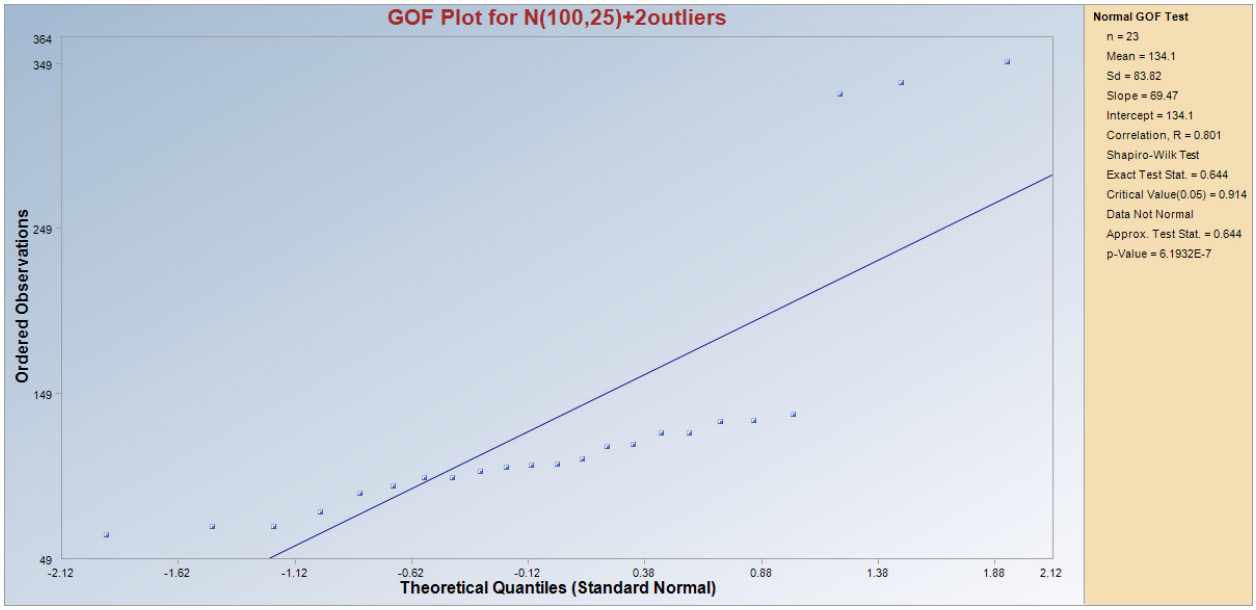

Figure 3. Formal Q-Q plot with Shapiro-Wilk Test Statistic and Indentifying Outliers Present in the Data Set.

Dixon Test Results on Data Set Used in Figure 3

Dixon's Outlier Test for N(100,25)+3outliers		
Mean	134.1	
Standard Deviation	83.82	
Number of data = 23		
10% critical value: 0.374		
5% critical value: 0.421		
1% critical value: 0.505		
1. 349.8921 is a Potential Outlier (Upper Tail)		
Test Statistic: 0.070		
For 10% significance level, 349.8921 is not an outlier.		
For 5% significance level, 349.8921 is not an outlier.		
For 1% significance level, 349.8921 is not an outlier.		
2. 63.1259374598842 is a Potential Outlier (Lower Tail)		
Test Statistic: 0.019		
For 10% significance level, 63.1259374598842 is not an outlier.		
For 5% significance level, 63.1259374598842 is not an outlier.		
For 1% significance level, 63.1259374598842 is not an outlier.		

Due to masking, the Dixon test failed to identify the 3 outliers present in the data set. The Rosner test cannot be used on this data set as the number of observations in this data set is less than 25. The graphical Q-Q plot shown in Figure 3 successfully identifies the 3 outliers (well-

separated from the main dominant data set) present in the data set. Today, modern robust statistical methods (e.g., Rousseeuw and Leroy, 1987; Rousseeuw and van Zomeren, 1990; Singh and Nocerino, 1995; Rousseeuw and van Driessen, 1999; Maronna, Marin and Yohai, 2006) are used to identify multiple outliers. Several commercial software packages (e.g., SAS, SPSS) are equipped with the robust estimation and outlier identification methods. Several of the robust methods are available in the EPA software package Scout 2008 (EPA 2009). Using the PROP influence function method of the Outlier Module of the Scout 2008 software (Singh, 1993), the three outliers present in the data set are identified. The formal Index Plot identifying all of the outliers present in the data set is shown in Figure 4. The details of these methods can be found in the User Guide for Scout 2008 (EPA 2009).

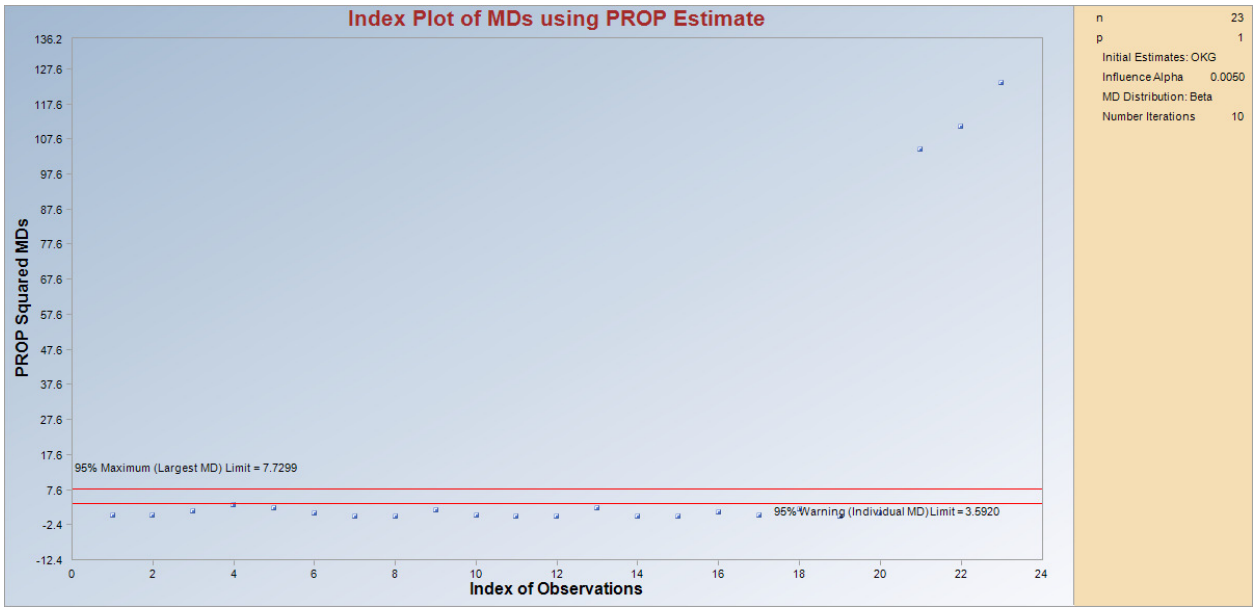

Figure 4. Formal Index Plot with Critical Value of the Test Statistic (Largest Mahalanobis Distance = Largest MD) Identifying the 3 Outliers Present in the Data Set

Use of Log-transformation was avoided

Statistical analysis including outlier tests and hypothesis testing approaches were performed on datasets in the original raw scale (non-transformed dataset) as the remediation and cleanup decisions need to be made using data and statistics (e.g., tolerance limits, prediction limits) in the original scale. Often, the use of a log-transformation tends to hide contamination by accommodating outlying observations (Singh et al., 1997; EPA, 2010b) as part of the dataset. For an example, an outlier in the raw scale may not appear to be an outlier in the transformed space (e.g., log-scale). This does not imply that the outlier (e.g., an elevated RBRA concentration in the original scale) identified in the original scale represents a clean unimpacted location and can be included in the computation of a BTV, estimated by upper prediction limit (UPL)/upper tolerance limit (UTL). Furthermore, since environmental decisions need be made based upon the values of statistics (e.g., UPL, t-test, Wilcoxon Rank Sum [WRS] test statistic) in the original scale, all transformed test statistics computed using log-transformation need to be back-transformed in the original scale. The transformation and back-transformation process yields statistics which suffer from an unknown amount of transformation bias.

Comparing DTL Data with Data Collected from the Three RBRAs

For select radionuclides (cobalt-60, cesium-137, plutonium-238, plutonium-239/240, strontium-90), surface soil datasets were collected from the 20 distant test locations (DTL). One of the objectives of this effort was to determine if the selected RBRAs represent unimpacted background locations. Specifically, surface soils data collected from the DTLs were compared with the surface soils data collected from the three RBRAs to determine if the sampling locations of the three RBRAs represent unimpacted background locations.

In this process, the first step was to identify high outlying observations potentially present in the combined surface soil dataset from the DTLs and the two formations/three RBRAs. Outliers (if any) were removed before comparing concentrations of DTLs and the three RBRAs. In addition to graphical Quantile-Quantile (Q-Q) plot, Rosner and Dixon outlier tests were used to identify high outliers potentially representing impacted locations. The details of these graphical and outlier tests can be found in ProUCL 4.00.05 Technical Guide (EPA, 2010b), *Data Quality Assessment: Statistical Methods for Practitioners* (EPA, 2006), and *Statistical Methods for Environmental Pollution Monitoring* (Gilbert, 1987).

In order to compare radionuclide activities of the three regions, DTLs and two formations, both graphical and statistical methods were used. Graphical methods used include side-by-side box plots and multiple Q-Q plots (details in the ProUCL 4.00.05 Technical Guide, 2010b), and statistical methods used include two-sample hypothesis tests and Oneway Analysis of Variance (ANOVA) test (EPA, 2000; EPA, 2002; EPA, 2006; EPA, 2009). Before computing BTV estimates based upon the established background data sets, Goodness-of-fit (GOF) tests were used to determine data distributions of the established background data sets. Depending upon the data distribution, appropriate parametric or nonparametric methods were used to compute estimates of BTVs. GOF tests as incorporated in ProUCL 4.1 (EPA, 2011) were used to determine distributions (normal, gamma, and lognormal) of the various datasets collected from the DTLs and three RBRAs. All GOF test statistics are supplemented with Q-Q plots which are commonly used to assess data distributions. The detailed description of the GOF tests, two-sample hypothesis tests, and Oneway ANOVA test used in this Background Evaluation Report are given in the ProUCL 4.00.05 Technical Guide (EPA 2010b).

Specifically, the two sample parametric t-tests (when applicable), nonparametric WRS or Wilcoxon Mann-Whitney (WMW) tests, Quantile tests, and Gehan tests were used to compare radionuclide concentrations of the: two formations, two RBRAs of the Chatsworth formation, and surface and subsurface soils of the three RBRAs and two formations. In some cases, for the verification of results and conclusions, more than one statistical method was used on the same dataset. Depending upon the data distribution, parametric and nonparametric methods were used to compare concentrations of two or more populations. Appropriate parametric (e.g., Student's t-test) or nonparametric (e.g., WMW) two-sample hypothesis tests were used to compare activities of two populations (e.g., DTLs versus Santa Susana, Santa Susana versus Chatsworth). For datasets with NDs, the Gehan test was used to compare concentrations of two populations. Appropriate parametric or nonparametric (e.g., Kruskal-Wallis test) Oneway ANOVA tests were used to compare concentrations of three (e.g., DTLs, Santa Susana, and Chatsworth) or more (e.g., DTLs versus three RBRAs) groups. Once it was determined that the concentrations of the selected RBRAs are comparable to those of DTLs, it is concluded that the selected RBRAs

(without outlying locations) indeed represent defensible site-specific unimpacted background locations.

Comparing Concentrations/Activities of Six Individual Datasets: Surface and Subsurface Soils of the Three RBRAs

Defensible background datasets free of outliers (potentially representing impacted locations) were established for the various radionuclides using the surface and subsurface soil data collected from the three RBRAs. Datasets collected from the various areas (e.g., surface and subsurface soils of three RBRAs) were compared to establish statistically/numerically different background populations. Separate BTV estimates were computed for each population (e.g., two different populations represented by surface and subsurface soils).

- Using the ProUCL 1.1 software, outliers were identified first in the combined dataset collected from the surface and subsurface soils of the three RBRAs. All identified outliers (if any) were removed from the combined dataset. It is noted that for most of the radionuclides, the datasets from the three RBRAs are fairly consistent and tight (with small variations) and do not consist of any outliers. These observations guided the Project Team to use more appropriate upper simultaneous limits (USL) to estimate BTVs (see Section 3.0 for details).

- To determine how many BTVs would need to be computed for each radionuclide, data from the six datasets (surface and subsurface soils of each of the three RBRAs) were compared using the hypothesis and ANOVA tests as described in the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (EPA, 2000) and EPA guidance documents (2002; 2006; 2010b).

- Surface and subsurface activities of the two formations were compared first.

    - Graphical and two sample hypothesis tests (t-test, WMW test, and Gehan test) were used to perform these comparisons.

    - If surface and subsurface activities are determined to be comparable, the Project Team may decide to use a single BTV estimate based upon the combined surface and subsurface dataset. The details of the methods used to compute estimates of BTVs are summarized in Section 3.0 of this appendix.

- If radionuclide activities in surface and subsurface soils of the two formations are determined to be different, concentrations of the three RBRAs were compared separately for surface and subsurface soils.

    - Graphical displays (box plots and Q-Q plots) and Oneway ANOVA tests were used to perform these comparisons.

    - If radionuclide activities in surface/subsurface soils of the three RBRAs are determined to be not comparable, radionuclide activities in surface/subsurface soils of Lang Ranch and Rocky Peak RBRAs of Chatsworth were compared.

- If radionuclide activities in surface/subsurface soils of Lang Ranch and Rocky Peak are comparable, then one BTV estimate may be used for the surface/subsurface soils of the Chatsworth Formation (combined Lang Ranch and Rocky Peak data), and one BTV estimate may be used for the surface/subsurface soils of the Santa Susana Formation (Bridle Path).

- In case radionuclide activities in the six datasets are not comparable, six BTV estimates may be used, one for each dataset.

3.0 STATISTICAL METHODS USED TO ESTIMATE BACKGROUND THRESHOLD VALUES BASED UPON ESTABLISHED BACKGROUND DATASETS WITH AND WITHOUT NONDETECT OBSERVATIONS

This section details the statistical methods and limits that have been used to estimate BTVs for the various radionuclides based upon well-established background datasets collected from the three background reference areas (Bridle Path, Lang Ranch, and Rocky Peak) of the two formations: Santa Susana and Chatsworth. Methods used to establish background datasets were described in Section 2.0.

Typically, in background evaluation and comparison studies, site-specific background level contaminant concentrations (data) are used to compare on-site concentrations with background level contaminant concentrations. These BTVs are estimated based upon well-established background datasets. Methods used to establish background datasets are summarized in Section 2.0 of this appendix. An onsite observation exceeding a BTV may be viewed as potentially coming from an impacted site area not belonging to the background population.

Let $x_1, x_2, \ldots, x_n$ represent a well-established background dataset of size n for a contaminant (e.g., cesium-137) collected randomly from some site-specific background/reference area. The objective is to estimate a BTV based upon this dataset. It should be pointed out that BTV estimates based upon parametric methods (e.g., normal, gamma distribution based) account for the variability present in the dataset; whereas nonparametric limits are based upon order statistics and, therefore, do not take data variability into consideration. It is also noted that nonparametric estimates of BTVs based upon order statistics (e.g., largest value, second largest value) tend to under estimate the background statistics, especially when the sample size is small (e.g., less than 60 samples). For an example, a nonparametric UTL95-95 or USL95 may not provide desired/specified confidence coefficient 0.95.

The sample mean and sample standard deviation are given as follows:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{n-1}}$$

The sample values are arranged in ascending order. The resulting ordered sample (called ordered statistics) is denoted by $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$. The ordered statistics are used as nonparametric

estimates of upper percentiles, UPLs, UTLs, and USLs. Let $y_i = \ln(x_i)$; $i = 1, 2, \ldots, n$, then $\bar{y}$ and s_y represent the mean and standard deviation (*sd*) of the log-transformed data. Once the data distribution of a background dataset has been determined, one can use parametric or nonparametric statistical methods to compute estimates of BTVs based upon the background dataset. Depending upon the sample size, data distribution and data variability, one of the following upper limits (EPA, 2011 [ProUCL 4.1]; Scout, 2008 [Version 1.0]; Singh and Nocerino, 1997) can be used to estimate BTVs.

Based upon well-established datasets collected from the three RBRAs of the two formations, the following limits have been computed:

- Upper percentiles
- UPLs
- UTLs
- USLs

Choosing a Confidence Coefficient (CC)

- Higher statistical limits are associated with higher levels of the confidence coefficient (CC). For an example, a 95% UPL is larger than a 90% UPL.

- Higher values of CC (e.g., 99%) tend to decrease the power of a test resulting in higher number of false negatives - dismissing contamination when present.
  Therefore, CC should not be set higher than necessary.

- Smaller values of CC (e.g., 0.80) tend to result in higher number of false positives (e.g., declaring contamination when not present).

- In most practical applications, choice of 95% CC provides a good compromise between confidence and power.

Sample Size

- Smaller sample sizes (e.g., fewer than 30 samples) tend to yield estimates with higher variabilities; which in turn result in higher values of USLs, UTLs and UPLs.

- Higher level of uncertainty in a background dataset (e.g., due to a smaller background dataset) tends to dismiss contamination as representing background conditions (results in more false negatives, i.e., identifying a location that may be dirty as background).

Parametric and Nonparametric Statistical Limits used to Estimate BTVs Based Upon Data without Nondetect Observations

In this background study report, 95% CC has been used to compute various upper limits to estimate BTVs, also, UTLs have been computed for the 95th percentile (coverage coefficient = 0.95). Depending upon the data distribution, both parametric and nonparametric methods have

been used to estimate BTVs. For parametric methods, BTV estimates have been computed for normal, lognormal, and gamma distributions.

Upper Percentiles to Estimate BTV: Based upon an established background (e.g., RBRA) dataset, the 95^{th} percentile ($x_{0.95}$) represents an estimate of the 95^{th} percentile of the background population. It is expected that 95% of values coming from the background population will be less than or equal to $x_{0.95}$. By definition, about 5% of observations from the background dataset will exceed $x_{0.95}$.

If the distributions of the site data and the background data are comparable, then an observation coming from a population (e.g., site) comparable to that of the background population should lie at or below the p*100% percentile, with probability p.

Normal distribution based $p*100^{th}$ sample percentile is given by the following statement:

$$\hat{x}_p = \overline{x} + sz_p$$

Here z_p is the $p*100^{th}$ percentile of a standard normal, N (0, 1) distribution, which means that the area (under the standard normal curve) to the left of z_p is p.

Interpreting 95% Percentile

- If one uses 95% percentile to estimate BTV, at least 5% of the on-site observations with concentrations comparble to background will be determined as not belonging to the background population even when they actually come from the background population.

- If an on-site value (x_{onsite}) exceeds background $x_{0.95}$, it may be concluded that x_{onsite} does not belong to the background population.

- The use of 95% percentile to estimate BTV potentially may lead to a higher number of false positives resulting in unnecessary cleanup (i.e., determining a clean on-site location comparable to background as dirty).

Upper Prediction Limits and Upper Tolerance Limits to Estimate BTV
Unlike upper percentiles, UPLs and UTLs provide predictive setup for future observations.

Upper Prediction Limits
UPLs can be computed for 1 or more future observations (e.g., future on-site values). Let UPL_k 95 represent a 95% UPL for *k (≥1)* future observations. A UPL_k 95 is designed to provide coverage for *k* future observations with CC 0.95.

95% Upper Prediction Limit ($UPL_1$95): The $UPL_1$95 is based upon a background dataset is designed to compare a *single future* observation with $UPL_1$95. We are 95% sure that a *"single"* future value from the background population will be less than or equal to $UPL_1$95 with a CC of 0.95. If an on-site value, $x_{onsite} < UPL_1$95, it is interpreted that x_{onsite} (=future value) comes from the background population with a CC of 0.95. A $UPL_1$95 is not meant to perform more than 1 future comparison.

Normal distribution: A normal distribution based $UPL_1(1-\alpha)100\%$ for a *single future* observation is given by the following probability statement:

$$P\left(x_0 \le \overline{x} + t_{((1-\alpha),(n-1))}s\sqrt{1+\frac{1}{n}}\right) = 1-\alpha \text{ , with } UPL_1 95 = \left(\overline{x} + t_{((1-0.05),n-1)}s\sqrt{1+\frac{1}{n}}\right)$$

Here $t_{((1-\alpha),(n-1))}$ is a critical value from Student's t-distribution with (n–1) degrees of freedom.

Interpreting $UPL_1 95$

- An on-site value (x_{onsite}) exceeding $UPL_1 95$ potentially represents a value not belonging to the background population.

- UPLs are useful when a background dataset is of smaller size (e.g., fewer than 20 to 30 samples); and/or a few and known number of future observations are to be compared with a UPL.

Improper Use of $UPL_1 95$ to Perform Many Future Comparisons

In practice, users tend to use $UPL_1 95$ for many future comparisons which results in a higher number of false postives (observations declared contaminated when in fact they are clean). When k future comparisons are made with a UPL_1, some of those future observations will exceed UPL_1 just by chance, each with probability 0.05. For proper comparison, UPLs need to be computed according to the number of comparisons that will be performed.

In order to achieve the specified false rejection rate of 0.05, we need to take the number of future comparisons into account; for a false positive error rate, α, UPL_k for k future observations should be used to estimate the BTV when k comparisons are to be performed with the BTV.

If many (as is the case in the present study) independent onsite comparisons (e.g., plutonium-238 activity from 30 on-site locations) are made with the same UPL_1, each on-site value may exceed that UPL_1 with probability 0.05 just by chance. The overall probability of at least one of those 30 comparisons being significant (exceeding BTV) just by chance is:

$\alpha_{actual} = 1-(1-\alpha)^k = 1 - 0.95^{30} \sim 1-0.21 = 0.79$ (false positive rate).

This means that the probability (overall false positive rate) is 0.79 that at least one of the 30 on-site locations will be considered contaminated even when they are comparable to background.

Similar arguments hold when multiple (=*m)* contaminants are analyzed, and status (clean or impacted) of an on-site location is determined based upon *m* comparisons (one for each analyte).

UPL_k for k Future Comparisons

95% Upper Prediction Limit ($UPL_k 95$) with k≥1 is designed to compare *k future* observations with $UPL_k 95$. We are 95% sure that *"k"* future values from the background population will be less than or equal to $UPL_k 95$ with a CC of 0.95. A $UPL_k 95$ is meant to

perform *k* future comparisons. A UPL_k uses an appropriate critical value (based upon Bonferroni inequality) to accommodate k future observations/comparisons. These UPLs statisfy the relationship: $UPL_1 \leq UPL_2 \leq UPL_3 \leq \ldots \leq UPL_k \ldots$

Normal distribution based UPL_3 for three future observations: $x_{01,} x_{02}, x_{03}$ is given by:

$$P\left(x_{01}, x_{02}, x_{03} \leq \overline{x} + t_{((1-\alpha/3),n-1)} s\sqrt{1+\frac{1}{n}} \right) = 1-\alpha \text{ With } UPL_3 95 = \left(\overline{x} + t_{((1-0.05/3),n-1)} s\sqrt{1+\frac{1}{n}} \right)$$

Normal distribution based UPL_k for k future observations is given by:

$$P\left(x_{01}, x_{02}, ..., x_{0k} \leq \overline{x} + t_{((1-\alpha/k),n-1)} s\sqrt{1+\frac{1}{n}} \right) = 1-\alpha \text{ With } UPL_k 95 = \left(\overline{x} + t_{((1-0.05/k),n-1)} s\sqrt{1+\frac{1}{n}} \right)$$

Upper Tolerance Limit (UTL): A UTL *(1-α)-p* (e.g., UTL95-95) based upon an established background dataset represents that limit such that *p*% of the sampled data will be less than or equal to UTL with a CC of (1-α). It is expected that p% of the observations belonging to the background population will be less than or equal to UTL with CC of (1-α). A UTL *(1-α)-p* represents a *(1-α) 100%* upper confidence limit for the p^{th} percentile of the underlying background population.

A UTL95-95 represents that statistic such that 95% observations from the target population (e.g., background) will be less than or equal to UTL95-95 with a CC of 0.95. UTL95-95 is designed to simultaneously provide coverage for 95% of all potential observations (current and future) from the background population with a CC of 0.95. UTL95-95 can be used to perform many on-site comparisons.

Normal Distribution Based UTLs: For normally distributed datasets, an upper *(1-α)*100%* tolerance limit with tolerance or coverage coefficient = *p* (that is providing coverage to at least *p*100% proportion of observations) is given by the following statement:

$$UTL = \overline{x} + K * s$$

Here, $K = K(n, \alpha, p)$ is the tolerance factor and depends upon the sample size, *n*, confidence coefficient = (1-*α*), and the coverage proportion = *p*. The values of the tolerance factor, *K*, have been tabulated extensively in the various statistical books (Hahn and Meeker, 1991). Those *K* values are based upon non-central t-distributions. Also, some large sample approximations (Natrella, 1963) are available to compute the *K* values for one-sided tolerance intervals (same for both UTLs and lower tolerance limit). The approximate value of *K* is also a function of the sample size, *n*, coverage coefficient, *p*, and the CC (1-*α*). In the ProUCL 4.1 software package, the values of *K* for samples of sizes ≤ 30, as given in Hahn and Meeker (1991) have been directly programmed. For sample sizes larger than 30, the large sample approximations, as given in Natrella (1963), have been used to compute the *K* values.

Interpreting UTL95-95

- UTL95-95 based upon a background dataset is that value which will be exceeded by all values potentially coming from the background population less than 5% of the time with confidence coefficient 0.95.

- For a UTL95-95, five exceedances per 100 comparisons (of background values) can result just by chance for an overall CC of 0.95; or 5% exceedances (in any number of comparisons) can occur just by chance with an overall CC of 0.95. Similarly, for UTL95-99, 1 exceedance per 100 comparisons can result just by chance for an overall CC of 0.95.

- Just like UPLs, a parametric UTL takes variability into account.

- When sample size is large (e.g., 500), UTL95-95 approaches the upper 95th percentile; UTL90-90 will approach the upper 90th percentile.

- Use of UTL95-95 is preferred to $UPL_1 95$ when the number of future comparisons is large and/or unknown.

Upper Simultaneous Limit (USL): A$(1-\alpha)100\%$ USL based upon an "established" background dataset is meant to provide coverage for all observations, x_i, $i = 1, 2, n$ simultaneously in the background dataset. A USL95 represents that statistic such that all observations from the "established" background dataset will be less than or equal to USL95 with a CC of 0.95. It is expected that observations coming from the background population will be less than or equal to USL95 with 95% CC. A USL95 can be used to perform many on-site comparisons.

Normal distribution based the two-sided $(1-\alpha)100\%$ simultaneous interval based upon the first order Bonferroni inequality (Singh and Nocerino, 1995; Singh and Nocerino, 1997) is given as follows:

$$P\left(\overline{x} - sd_{\alpha}^{b} \le x_i \le \overline{x} + sd_{\alpha}^{b}; i := 1, 2, ..., n\right) = 1 - \alpha .$$

Here, $(d_{\alpha}^{b})^2$ represents the critical value (obtained using the Bonferroni inequality) of the maximum Mahalanobis distance (Max MD) for α level of significance (Singh and Nocerino, 1995; Singh, 1996).

In order to maintain proper balance between false positives and false negatives, the use of USL95 was proposed to estimate BTVs based upon well-established background data sets free of outliers. It should be noted that a USL95 has a built-in outlier identification test (Wilks, 1963; Gnanadesikan, 1977; Barnett and Lewis, 1994; Singh and Nocerino, 1997). This means that observations lying above a USL95 potentially represent outliers not belonging to the main background population.

Inclusion of outliers in background data sets will yield inflated estimates of BTVs including USL95 values. The use of USL95 already addresses the issue of increased number of false

positives. To control the number of false negatives, it is recommended not to include moderate to extreme outliers in the computation of USL95. USL95 should be computed based upon a data set representing the main dominant background population; it is not desirable to accommodate a few outliers in the computation of USL95 resulting in inflated USL95.

Normal distribution based one-sided *(1 - α)100%* USL is given by:

$$P\left(x_i \le \overline{x} + sd_{2\alpha}^{b}; i := 1, 2, ..., n\right) = 1 - \alpha \ ;$$

$$\text{USL} = \overline{x} + d_{2\alpha}^{b} * s_x$$

Here $(d_{2\alpha}^{b})^2$ is the critical value of the Max MD for 2*α level of significance.

Interpreting USL95

In layman's terminology, a USL95 provides coverage to all observations (current and future) coming from the same background population with probability 0.95. Depending upon the data variability, some observations (current and future) will exceed USL95 with probability 0.05. The computation of USL95 depends upon the sample size, data mean and variability, and the critical value of the test statistic used. Sample values (e.g., maximum value) exceeding USL95 potentially represent extreme values and may not be considered as coming from the background population.

- A USL95 based upon an established background dataset represents that estimate of BTV such that *all* observations in the background dataset are less than or equal to USL95 with 95% CC.

- It is expected that observations coming from that background population will be less than or equal to USL95 with a 95% CC.

- A USL can be used when many (and/or unknown) future on-site observations need to be compared with BTV.

- For an analyte, the false positive error rate does not change with the number of comparisons, as USL95 is meant to perform many comparisons simultaneously. The use of USL95 is preferred when many comparisons need to be performed.

If BTVs are to be computed based upon the background data sets incorporating data variability, it is suggested to use USL95 as computed based upon the data set without using biased judgment about the data set and the computed statistics.

Other parametric (e.g., lognormal and gamma) and nonparametric limits used to estimate BTVs are described as follows.

Lognormal Distribution Based Upper Limits

Lognormal Percentiles: Lognormal p*100th percentile is given by the following statement:

$$\hat{x}_p = \exp(\overline{y} + s_y z_p)$$

Where, z_p is the upper p*100th percentile of a standard normal, N(0,1), distribution.

Lognormal UPLs: An upper (1 – α)*100% lognormal UPL is given as follows:

$$\text{UPL} = \exp(\overline{y} + t_{((1-\alpha),(n-1))} * s_y * \sqrt{(1+1/n)})$$

As before, $t_{((1-\alpha),(n-1))}$ represent the critical value from Student's t-distribution with (n–1) degrees of freedom.

Lognormal UTLs: The procedure to compute UTLs for lognormally distributed datasets is similar to that for normally distributed datasets. The sample mean, $\overline{y}$, and standard deviation (sd), s_y, of the log-transformed data are computed using a defensible unique background dataset without outliers. An upper (1 – α)*100% tolerance limit with tolerance or coverage coefficient = p (that is, providing coverage to at least p100% proportion of observations) is given by the following statement:

$$\text{UTL} = \exp(\overline{y} + K * s_y)$$

Just like the normal distribution, the UTL given above represents a (1-α)*100% confidence interval for the p^{th} percentile of the lognormal distribution. The critical value, K used to compute the UTL is the same as the one used to compute the normal UTL.

Lognormal USL: An upper (1 – α)*100% lognormal USL is given by the following equation:

$$\text{USL} = \exp(\overline{y} + d^b_{2\alpha} * s_y)$$

Here $(d^b_{2\alpha})^2$ is the critical value of the Max MD for 2*α level of significance as described earlier.

Gamma Distribution Based Upper Limits

Positively skewed environmental data can often be modeled by a gamma distribution (Singh et al., 2002). ProUCL 4.1 software has two goodness-of-fit tests (Anderson-Darling test and Kolmogorov-Smirnov test) to test for gamma distribution. UPLs and UTLs obtained using normal approximation to the gamma distribution (Krishnamoorthy et al., 2008) have been incorporated in ProUCL 4.1. Those approximations are based upon Wilson-Hilferty - WH (1931) and Hawkins-Wixley - HW (1986) approximations.

- According to WH approximation, the transformation, $Y = X^{1/3}$ follows an approximate normal distribution.

- According to HW approximation, the transformation, $Y = X^{1/4}$ follows an approximate normal distribution

In the following, $\overline{y}$ and s_y are the mean and standard deviation of the observations in the transformed scale (Y) defined above.

Gamma Distribution Percentiles: The details of computing percentiles of a gamma distribution, *G (k,θ)* can be found in the ProUCL 4.00.05 Technical Guide (2010b). Let $y_{(\alpha\ 2k)}$ represent the α*100th percentile of the chi-square distribution with 2*k* degrees of freedom. The α*100% percentile for a gamma distribution can be obtained by using the equation:

$$x_\alpha = y_\alpha * \theta/2$$

Gamma Distribution Based UPLs: Given a sample, $x_1, x_2, \ldots, x_n$ of size n from a gamma, G(k, θ) distribution, approximate (based upon WH and HW approximations) (1 – α)*100% upper prediction limits for a future observation from the same gamma distribution are given by:

$$\text{Wilson-Hilferty (WH) UPL} = \max\left(0, \left(\overline{y} + t_{((1-\alpha),(n-1))} * s_y * \sqrt{1 + 1/n}\right)^3\right)$$

$$\text{Hawkins-Wixley (HW) UPL} = \left(\overline{y} + t_{((1-\alpha),(n-1))} * s_y * \sqrt{1 + 1/n}\right)^4$$

$t_{((1-\alpha),(n-1))}$ is the critical value from Student's t-distribution with (*n*–1) degrees of freedom. The UPLs for the next k>1 (k future observation) can be obtained similarly and have been incorporated in ProUCL 4.1.

Gamma Distribution Based UTLs: Using the WH approximation, the gamma UTL based upon a sample of size n (in original scale, X), is given by:

$$UTL = \max\left(0, \left(\overline{y} + K * s_y\right)^3\right)$$

Similarly, using the HW approximation, the gamma UTL in original scale is given by:

$$UTL = \left(\overline{y} + K * s_y\right)^4$$

Here K represents the same cutoff value as used for normal and lognormal distributions.

Gamma Distribution Based USLs: Given a sample, $x_1, x_2, \ldots, x_n$ of size n from a gamma, G(k, θ) distribution, approximate (based upon WH and HW approximations) (1 – α)*100% upper simultaneous limit is given by:

$$\text{Wilson-Hilferty (WH) USL} = \max\left(0, \left(\overline{y} + d^b_{2\alpha} * s_y\right)^3\right)$$

Hawkins-Wixley (HW) USL = $\left(\overline{y} + d_{2\alpha}^{b} * s_{y}\right)^{4}$

Here $(d_{2\alpha}^{b})^{2}$ is the critical value of the Max MD for 2*α level of significance.

Nonparametric Upper Limits

When the background dataset does not follow a discernable distribution, nonparametric methods have been used to estimate BTVs as described below.

Upper Percentile: Details can be found in the ProUCL Technical Guide (EPA, 2010b). It is noted that the nonparametric 95% percentile ($x_{0.95}$) does not take variabilities of future observations coming from the background population into account.

Nonparametric Upper Prediction Limits: A one-sided nonparametric UPL is simple to compute and is given by the following m^{th} order statistic. One can use linear interpolation if the resulting number, m, given below does not represent a whole number (a positive integer).

$$\text{UPL} = X_{(m)}, \text{ where } m = (n + 1) * (1 - \alpha).$$

For example, for a nonparametric dataset of size 25, a 90% UPL is desired. Then m = (26*0.90) = 23.4. Thus, a 90% nonparametric UPL can be obtained by using the 23rd and the 24th ordered statistics and is given by the following equation:

$$\text{UPL} = X_{(23)} + 0.4 * (X_{(24)} - X_{(23)})$$

Nonparametric Upper Tolerance Limits: The computation of nonparametric UTLs is somewhat messy as it is based upon binomial cumulative probabilities and order statistics. Just like parametric UTLs, a nonparametric UTL is supposed to provide coverage to p*100% observations from the target population with a specified CC. It is noted that nonparametric UTLs (computed by order statistics) cannot exactly achieve the specified confidence coefficient (1-α). In most cases, only an approximate CC can be achieved by nonparametric UTLs. One has to be satisfied with the achievable CC, which is as close as possible to the specified CC of (1-α). Thus, an appropriate UTL is chosen which provides coverage for the p^{th} percentile as close as possible to the specified CC of (1-α). The details about the computation of nonparametric UTLs can be found in ProUCL 4.00.05 Technical Guide (2010b), Conover (1999), Gogolak et al. (1998), David and Nagaraja (2003), and Hahn and Meeker (1991).

Nonparametric Upper Simultaneous Limit: A nonparametric USL is estimated by the largest value of the dataset.

Statistical Limits used to Estimate BTVs Based Upon Dataset Consisting of Nondetect / Below Detection Limit Observations

For datasets consisting of NDs with multiple detection limits or minimum detectable concentrations (MDC), several estimation methods including the maximum likelihood (MLE) method, Kaplan-Meier (KM) method, bootstrap methods, regression on order statistics (ROS) methods, and substitution methods are available in ProUCL 4.1 (EPA, 2011) and Scout 2008 Version 1.0 (EPA, 2009) software packages. In this study, the preferred KM method (Helsel, 2005; Singh et al., 2006; Kaplan and Meier, 1958) has been used to compute estimates of BTVs.

The details of the various estimation methods including the KM method can be found in the ProUCL 4.00.05 Technical Guide (EPA, 2010b), Singh et al. (2006), and Helsel (2005). A brief description of the KM method and various limits used to estimate BTVs based upon the KM method is given as follows. ProUCL 4.1 and Scout software packages have been used to generate graphical displays and compute the various BTV estimates based upon datasets consisting of ND observations.

Nonparametric Kaplan-Meier (KM) Estimation Method

Practitioners (Helsel, 2005; Singh et al., 2006) and EPA guidance documents (2009) recommend the use of KM method when dealing with environmental datasets consisting of ND (values reported with U flags; values less than respective MDC) observations. For datasets with ND observations with multiple detection limits, the KM estimation method has been incorporated in ProUCL 4.1. A brief description of the KM method to estimate the population mean and standard deviation, and standard error (SE) of the mean is given as follows.

Let $x_1, x_2, ..., x_n$ (reported ND values or actual measurement) represent n data values obtained from samples collected from an area of concern (AOC), and let $x'_1 < x'_2, \ldots < x'_n$ denote the n' distinct values at which detects are observed. That is, n' ($\leq n$) represents distinct observed values in the collected dataset of size n. For $j = 1, \ldots, n'$, let m_j denote the number of detects at x'_j and let n_j denote the number of $xi \leq x'_j$. Also, let $x(1)$ denote the smallest x_i. Then:

$$\tilde{F}(x) = 1, \qquad x \leq x'_{n'}$$

$$\tilde{F}(x) = \prod_{j \ such \ that \ x'_j > x} \frac{n_j - m_j}{n_j}, \qquad x'_1 \leq x \leq x'_{n'}$$

$$\tilde{F}(x) = \tilde{F}(x'_1), \qquad x(1) \leq x \leq x'_1$$

$$\tilde{F}(x) = 0 \text{ or undefined}, \qquad 0 \leq x \leq x(1)$$

Note that in the last equality statement of $\tilde{F}(x)$ above, $\tilde{F}(x) = 0$ when $x(1)$ is a detect, and is undefined when $x(1)$ is a ND. The estimation of the population mean using the KM method is described as follows.

$$\hat{\mu} = \sum_{i=1}^{n'} x'_i [\tilde{F}(x'_i) - \tilde{F}(x'_{i-1})], \text{ with } x_0 = 0$$

An estimate of the SE of the mean is given by the following equation:

$$\hat{\sigma}^2_{SE} = \frac{n-k}{n-k-1} \sum_{i=1}^{n'-1} a_i^2 \frac{m_{i+1}}{n_{i+1}(n_{i+1} - m_{i+1})},$$

Where k = number of observation below the detection limit and

$$a_i = \sum_{j=1}^{i} (x'_{j+1} - x'_j)\tilde{F}(x'_j)\,,\ i\text{: =1, 2, …, } n'\text{-1.}$$

The KM method based mean is also denoted by $\hat{\mu} = \hat{\mu}_{KM}$. The KM method based estimate of the population variance is computed by using the following equation:

$\hat{\sigma}^2_{KM} = \hat{\mu}_{(x^2)-KM} - \left(\hat{\mu}_{(x)-KM}\right)^2$, where $\hat{\mu}_{(x)-KM}$ = KM mean of the data

$\hat{\mu}_{(x^2)-KM}$ = KM mean of the square of the data (second raw moment)

The literature referenced above suggests using the following equation to compute the UCL95 of the mean.

$\text{UCL95} = \hat{\mu} + t_{0.95,(n-1)}\sqrt{\hat{\sigma}^2_{SE}}$

Computation of KM Method Based Estimates of BTVs

The following limits (Helsel, 2005; Singh, Maichle, and Lee, 2006; EPA, 2010b; EPA, 2009) have been used to compute BTV estimates for datasets consisting of below detection limit (e.g., MDC) observations.

Percentiles: The p^{th} percentile based upon KM estimates (as incorporated in ProUCL 4.1) is given as follows.

$\hat{x}_p = \hat{\mu}_{KM} + z_p\sqrt{\hat{\sigma}^2_{KM}}$

$\hat{\mu}_{KM}$ = KM method based estimate of mean;

$\hat{\sigma}_{KM}$ = KM method based estimate of the population standard deviation; and

z_p is the p*100th percentile of a standard normal, N (0, 1) distribution, which means that the area (under the standard normal curve) to the left of z_p is p.

If the distributions of the site data and the background data are comparable/similar, then an observation coming from a population (e.g., site) similar to that of the background population should lie at or below the p*100% percentile, with probability p.

Upper Prediction Limit (UPL): For small samples (e.g., fewer than 30 samples), UPL can be computed using the critical value from the Student's t-distribution; and for large datasets, UPL can be computed using the critical values from normal distribution.

$$UPL = \hat{\mu}_{KM} + t_{((1-\alpha),(n-1))} * \hat{\sigma}_{KM}\sqrt{1+\frac{1}{n}}$$

$$UPL = \hat{\mu}_{KM} + z_{\alpha} * \hat{\sigma}_{KM}\sqrt{1+\frac{1}{n}}$$

$\hat{\mu}_{KM}$ = Kaplan Meier estimate of mean based upon data x_i, $i = 1, 2, \ldots, n$;

$\hat{\sigma}_{KM}$ = Kaplan Meier estimate of population standard deviation;

$t_{((1-\alpha),(n-1))} = (1-\alpha)^{th}$ critical value from t-distribution with degrees of freedom of (n-1);

$z_\alpha = \alpha^{th}$ critical value from standard normal distribution

Upper Tolerance Limit (UTL): Just like Student's t-statistic-based UPL, the use of the following equation has been used to compute upper (1-α)*100% tolerance limits with tolerance or coverage coefficient = p (that is providing coverage to at least p*100% of observations):

$$\text{UTL} = \hat{\mu}_{KM} + K * \sqrt{\hat{\sigma}^2_{KM}}$$

As before, K = K (n, α, p) is the tolerance factor and depends upon the sample size, n, CC = (1-α), and the coverage proportion = p. The K critical values are based upon non-central t-distribution, and have been tabulated extensively in the statistical literature (Hahn and Meeker, 1991).

Upper Simultaneous Limit (USL): An upper (1-α)*100% KM method based USL is given by the following equation:

$$\text{USL} = \hat{\mu}_{KM} + d^b_{2\alpha} * \hat{\sigma}_{KM}$$

Where $(d^b_{2\alpha})^2$ is the critical value of the Max (MDs) for 2*α level of significance.

4.0 ALTERNATIVE METHODS TO COMPUTE STATISTICS OF INTEREST: MEAN, STANDARD DEVIATION, UCLS, UTLS, USLS, UPLS BASED UPON RADIONUCLIDE DATASETS CONSISTING OF NONDETECTS

In the absence of the well-established statistical literature and theoretical guidance on how to compute various statistics of interest based upon radionuclide datasets consisting of NDs (U-flagged values), the available established statistical methods as described in the previous sections were used to compute BTV estimates for the SSFL Background Study. Historically, due to the complexity of statistical methods and lack of computer programs to compute decision making statistics (e.g., UPLs, UTLs, USLs) based upon datasets consisting of NDs (e.g., represented by negative radionuclide results), the use of some rule-of-thumb methods (without theoretical backup and detailed thought process) have been suggested based upon professional judgement/opinion in the various environmental documents (EPA, 2000). Those suggestions/statements have been made due to the lack of the availability of computer programs (at the time) to properly analyze radionulcide datasets with NDs represented by positive or negative results.

Team Decision to Compute BTV Estimates Based upon Dataset with Nondetects (U Flag)
At the start of the SSFL Background Study Project, the members of the Project Team including: Nicole Moutoux, Gregg Dempsey, Daniel Stralka, Jeffrey Bigelow, and Anita Singh decided to treat NDs as NDs, and use the available well-established statistical methods (Helsel, 2005; Singh

et al., 2006) to statistically analyse datasets consisting of ND observations. A considerable amount of effort was spent by the data validators to properly report NDs with U flags. In order to compute accurate and reliable estimates of BTVs, it is important to take this information (ND status with U flags) into account in the computation of various decision making statistics (e.g., UTLs, USLs). The statistical methods used to compute estimates of BTVs for the various radionuclides (and summarized above) properly account for the ND status of the reported radionuclide concentrations. If one has to ignore the non-detect status of concentrations reported with U flag, then there is no point in spending so much time during the data validation process to tag concentrations with U flag.

It was also decided by the Project Team that for radionuclides with no detected values or only a few detects (less than 5), BTVs will be estimated by their respective largest ND values. The largest ND still represents a ND, therefore its use as an estimate of BTV should not be considered as an overestimate of the BTV potentially resulting in false negatives.

Alternative Approach to Handle Datasets With NDs

Some technical stakeholders believe that radionuclide data consisting of NDs (positive as well as negative results) should be treated as detected data. They seem to suggest that one should ignore the ND status of radionuclide concentrations and their detection limits/MDCs. All detected as well as ND values should be treated equally in the computation of various statistics of interest including BTV estimates. They do not acknowledge the fact that in practice concentrations cannot be negative.

It is noted that radionuclides are reported as negative values due to background noise and limitation of the instruments. Some technical stakeholders argue that the bias in radionuclide concentrations cancels out as the uncertainty is associated with each result; sometime positive and sometime negative. However, the background noise (uncertainty in activity of a specific sample) cancels out for a specific single sample when the same sample is analyzed many times but not for all samples combined together. It is noted that the intensity of background noise (uncertainty in a result) depends upon activity/concentration of the samples that are being analyzed.

Note: The appropriateness of the approach proposed by these technical stakeholders is not clear and well-established. Specifically, it is not clear if the BTV estimates based upon the proposed approach provide the desired confidence associated with the various limits (e.g., UTL95-95, USL95) used to estimate BTVs. The proposed approach requires further investigation perhaps via monte carlo simulation experiments.

The Project Team for the SSFL Background Study believes that the altermative technical stakeholder approach results in biased estimates of BTVs and other statistics of interest. In the absence of the published literature illustrating the proposed approach, the author of this report and other members of the Project Team prefer to use the documented and well-researched statistical methods to deal with radionuclide datasets with ND observations.

The Project Team believes and recommends that instead of computing distorted BTV estimates by including NDs directly in the computation of upper limits, the status of ND observations

should be taken into consideration by using appropriate statistical methods designed to deal with datasets with NDs.

Some Observations

- It needs to be emphasized that if all or most (e.g., greater than 95%-99%) of results in a dataset are reported as NDs (uncensored or censored), then decision statistics (e.g., UCL, UPL, UTL, USL) based upon such a dataset may also be considered as NDs.

    - The approach which ignores the status of results reported as NDs tends to yield BTV estimates higher than what they should have been, especially when the majority of results are reported as NDs (see examples below).
    - By treating NDs as detects, the valueable information about NDs (U flagged data) is lost and the resulting statistics (e.g., USL, UTL) used to estimate BTVs and UCL95 used to estimate the exposure point concentration (EPC) term may be misleading and/or incorrect.

- Negative NDs when treated as detects increase the dataset variability; which in turn may result in inflated estimates of BTVs and EPC term; this is especially true when % of NDs is high.

    - Treating negative NDs as detects may result in a dataset without a discernable distribution. BTV estimates for such datasets are typically estimated by the higher order statistics (e.g., the largest, the second largest,..) which do not take the data variability into consideration. Depending upon sample size, nonprametric upper limits may overestimate the BTV estimate (e.g., for larger sample sizes greater than 100); or may not provide the specified (e.g., 95%) coverage (that is the BTV estimates are lower than what they should have been), especially when dataset is of smaller size (e.g., less than 60).

- By treating NDs as detects – the resulting statistics (e.g., estimates of BTVs) may become larger than the largest ND (examples below); and it is not clear how to interpret such a decision statistic - a detect or a ND.

Examples Illustrating the Issues Listed Above

Some issues associated with the technical stakeholders approach to compute decision statistics (e.g., UCLs, UTLs, USLs) are discussed in the following examples. For comparison sake, statistics of interest are computed using the two approaches: 1) treating NDs as detects as suggested by technical stakeholders; and 2) KM method which treats NDs as NDs. Real data sets collected from DOE sites considered in Examples 2-4 were provided by Dr. Beal of SAIC (a colleague of Dr. Rucker).

Example 1. Consider a small cobalt-60 (Co-60) dataset of size 14 from a real site. About 29% results of Co-60 are reported as NDs, and all NDs are positive. The dataset of size 14 obtained by treating all NDs as detects follows a normal (and also lognormal, gamma) distribution. Data are: 0.796 (U), 0.6 (U), 4.3, 1.1, 3.1, 2.5 (U), 0.1, 1.6, 1.2, 4, 3.9, 0.2, 1.04 (U), and 1.7.

	C-60	
Number of Observations	14	
Number of Missing Values	0	
Number of Detects	10	
Number of Non-Detects	4	
Percentage of Non-Detects	28.57%	
Minimum Non-Detect Value	0.6	
Maximum Non-Detect Value	2.5	
Minimum Observed Detected Value	0.1	
Maximum Observed Detected Value	4.3	
Mean of Detected Values	2.12	
Median of Detected Values	1.65	

Normal distribution based BTV estimates (treating NDs as detects) are shown in Figure 5 and the KM method based BTV estimates (treating NDs as NDs) are shown in Figure 6.

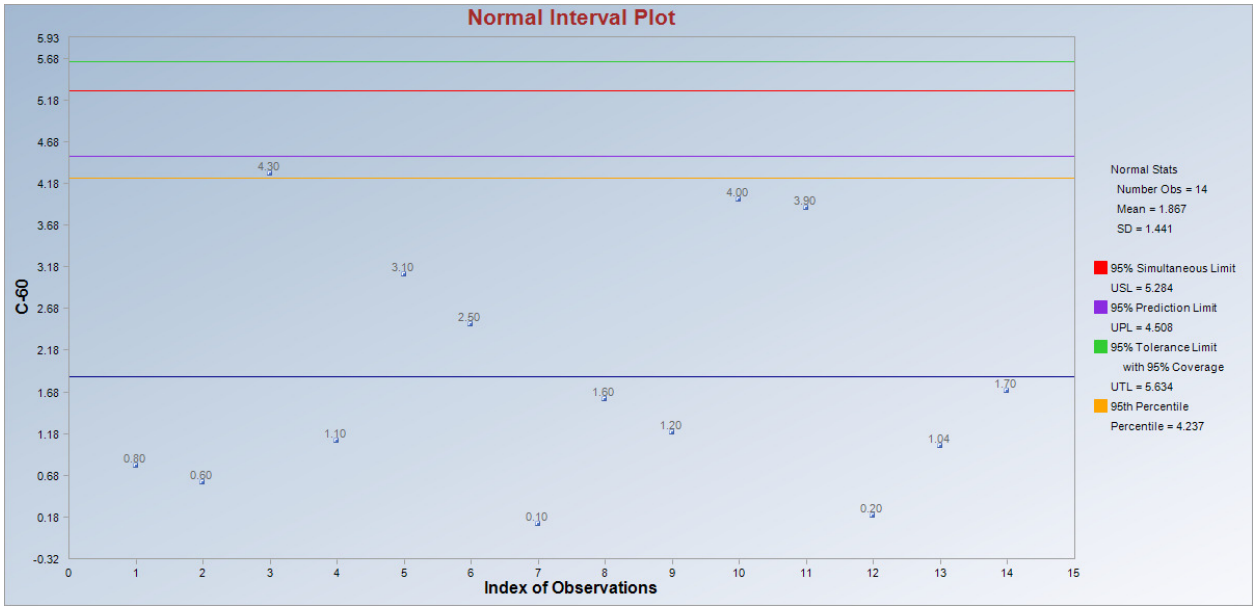

Figure 5 - Normal Distribution Based BTV Estimates for Co-60 in Site 1

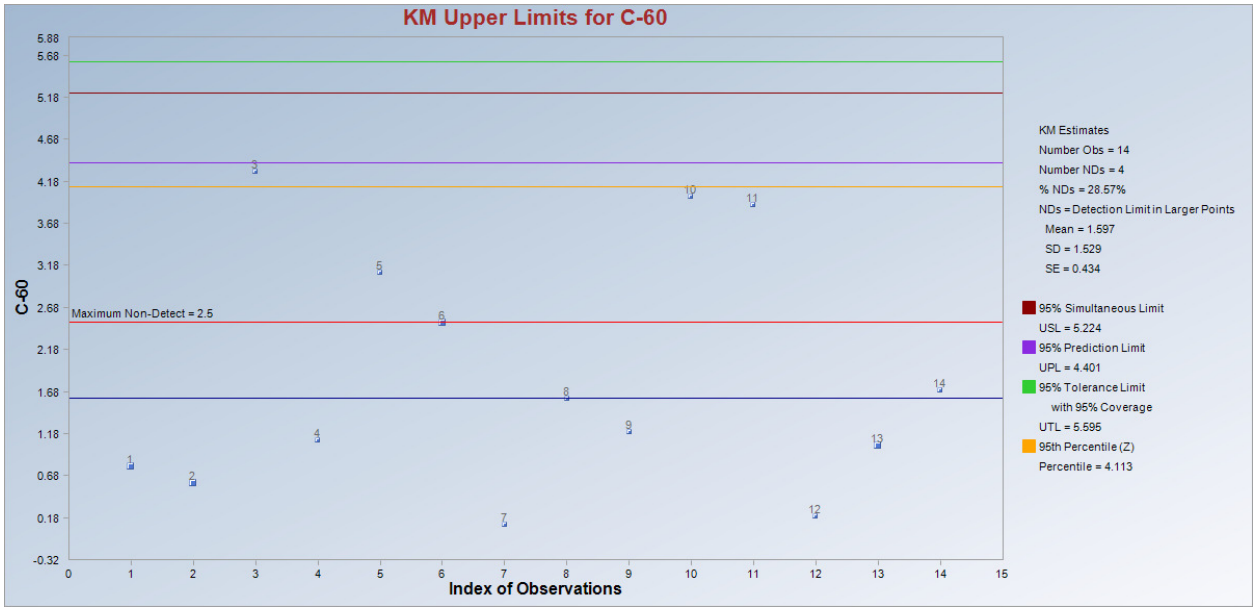

Figure 6 - KM Method Based BTV Estimates for Co-60 from Site 1

For this small dataset of size 14 with all positive NDs, the differences in BTV estimates obtained using the two approaches are not significantly large.

Example 3. Consider the cesium-137 (Cs-137) dataset of size 14 from a real site. All activity results are reported as NDs (some positive and some negative). Since all results are NDs - mean, UCL, UTL, USL should also be considered as NDs. When treating NDs as detects, the presence of negative and positive values increases the data variability, which in turn will result in higher estimates of BTVs.

	Cs-137
Number of Observations	14
Number of Missing Values	0
Number of Detects	0
Number of Non-Detects	14
Percentage of Non-Detects	100%
Minimum Non-Detect Value	-3.47
Maximum Non-Detect Value	6.36

Using the approach suggested by some technical stakeholders, treating ND Cs-137 results as detects, the Cs-137 dataset follows a normal distribution and normal distribution BTV estimates are shown below:

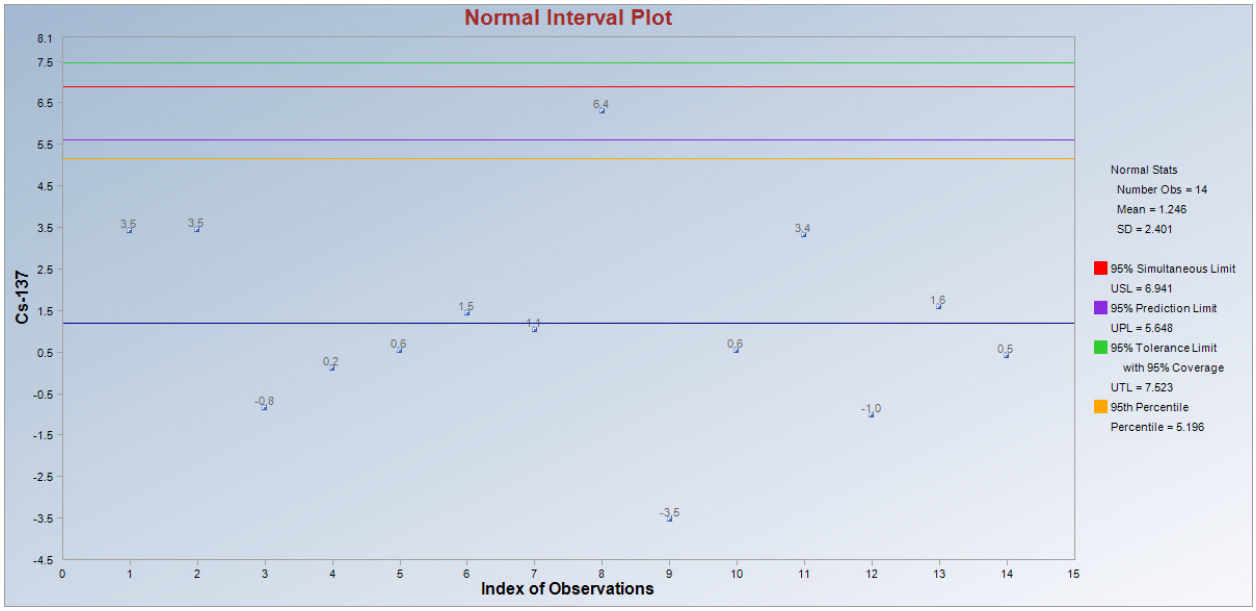

Figure 7 - Cs-137 Dataset of Size 14 with All Results Reported as NDs

Normal distribution based USL95= 6.941, and UTL95-95 = 7.52. Both of these values exceed the largest ND of 6.36.

It is not clear whether USL95 and UTL95-95 represent nondetects or detects.

Example 4. Just like Cs-137, all Co-60 results from this site are reported as NDs. Since all results are NDs, mean, UCL, UTL, USL should also be considered as NDs. Perhaps the BTV estimate may be considered as the largest non-detect value of 4.34.

	C-60
Number of Observations	14
Number of Missing Values	0
Number of Detects	0
Number of Non-Detects	14
Percentage of Non-Detects	100%
Minimum Non-Detect Value	-3.6
Maximum Non-Detect Value	4.34

Treating ND Co-60 results as detects, the Co-60 dataset follows a normal distribution. Presence of negative results increased the dataset variability, which resulted in inflated estimates of BTVs. Normal distribution based BTV estimates are shown below in Figure 8.

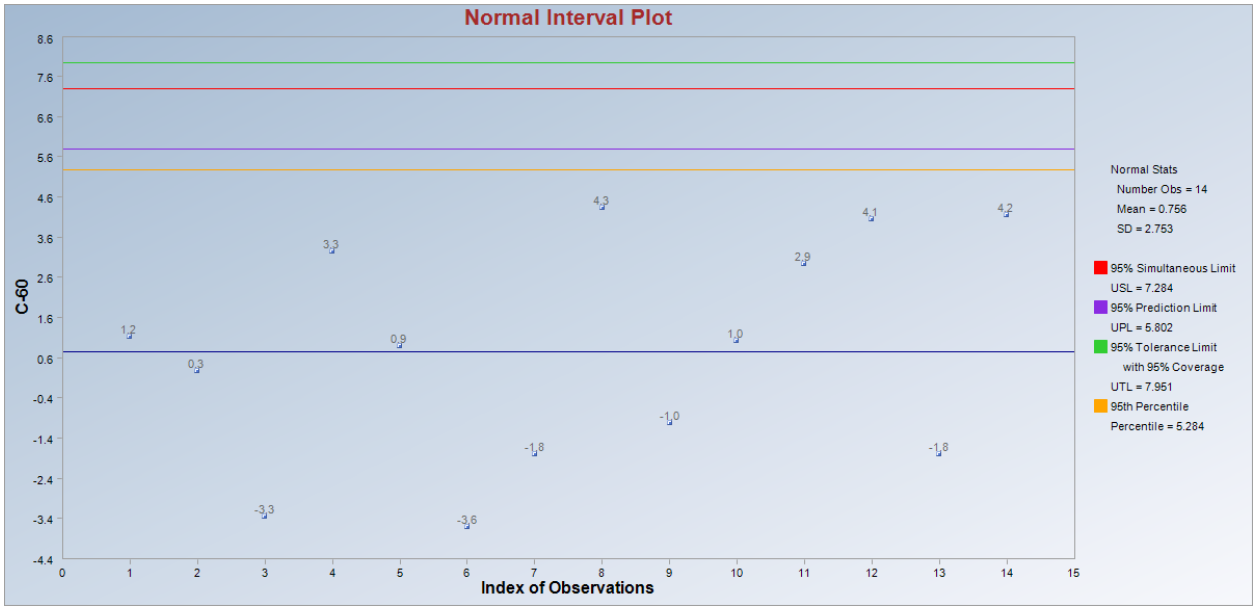

Figure 8 – C0-60 Dataset of Size 14 with All Results Reported as NDs

Normal distribution based USL95= 7.28, and UTL95-95 = 7.95. Both of these values exceed the largest ND of 4.34. The use of inflated USL95 to estimate BTV may result in a larger number of false negatives.

Also, it is not clear whether USL95 and UTL95-95 represent NDs or detects.

Example 5. Consider the uranium-232 (U-232) dataset collected from the reference areas of the two formations selected for the SSFL Background Study.

The datasets consists of only 1 (out of 149) detect = 0.417. Therefore, an estimate of BTV is also considered a ND.

Now if one treats NDs as detects – the detected value 0.417 represents an outlier as shown in Figure 9 below. The outlier is removed; the resulting dataset follows a normal distribution (Figure 10). Normal distribution based BTV estimates (without outlier) are shown in Figure 11.

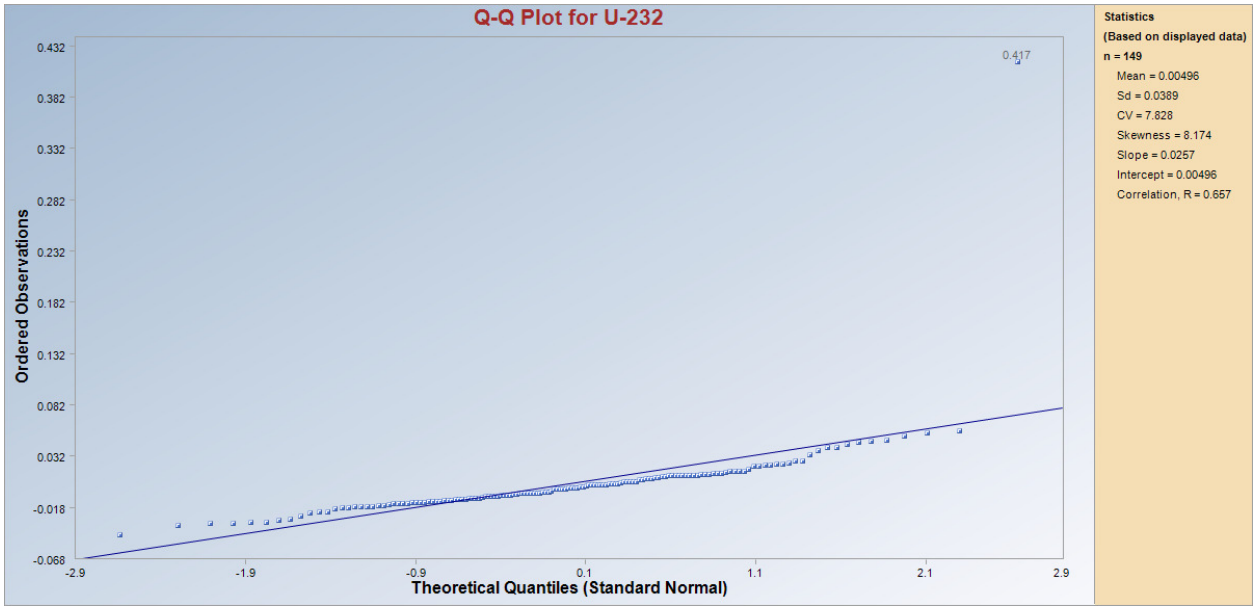

Figure 9 - Q-Q Plot Showing Observation 0.417 as Outlying

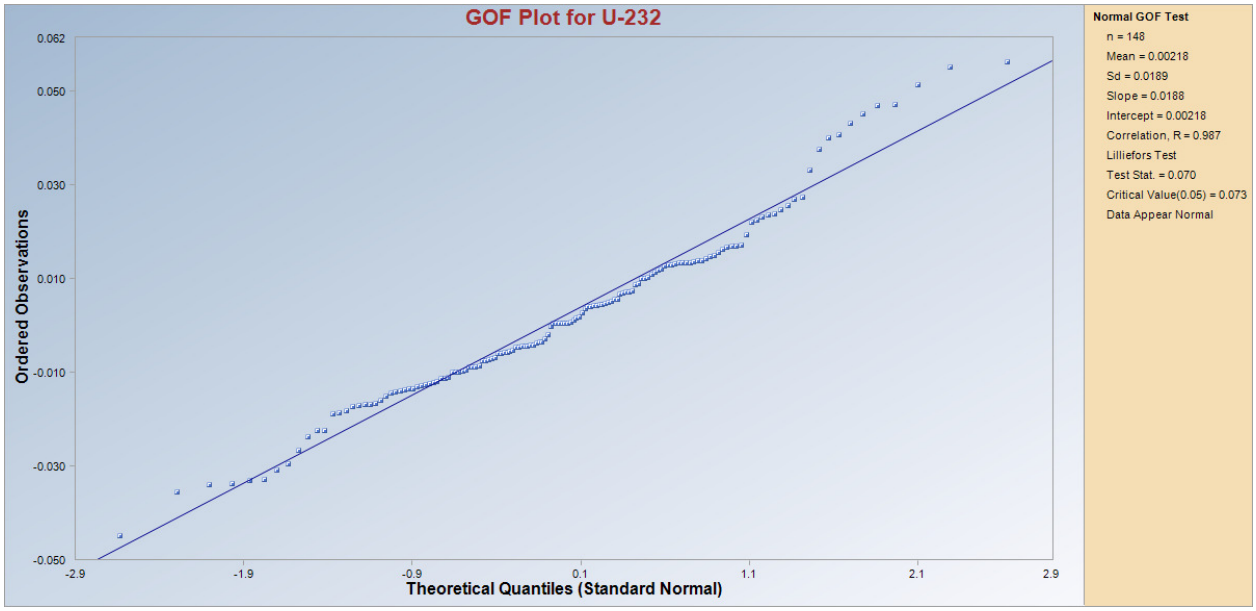

Figure 10 - U-232 Follows a Normal Distribution (without outlier) Treating NDs as Detects

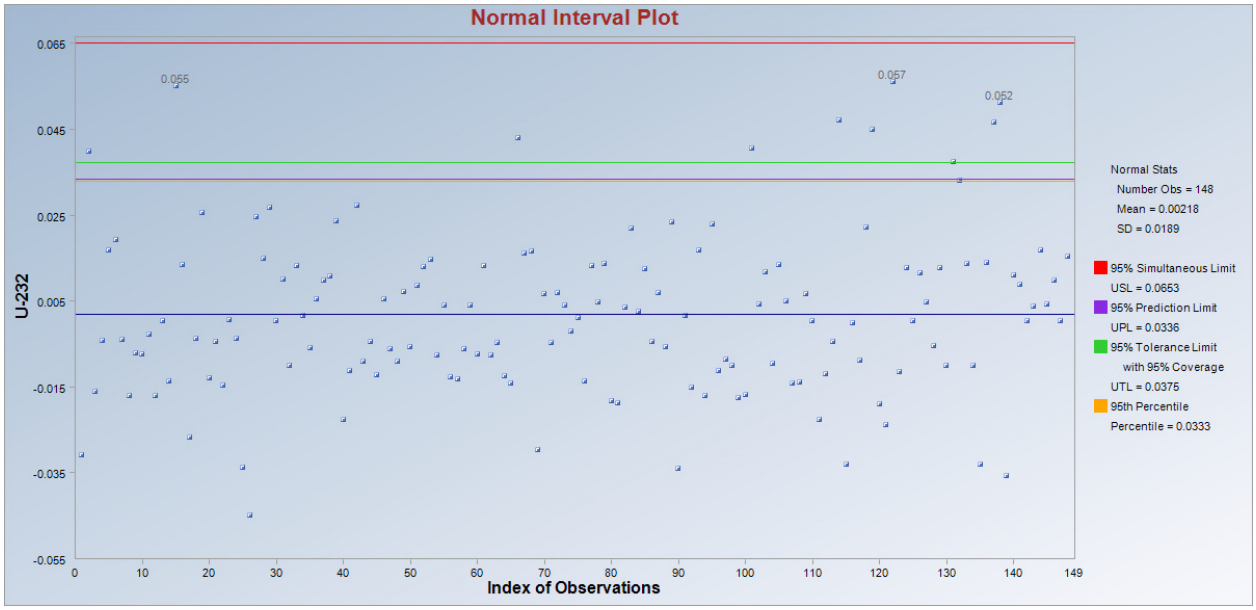

Figure 11 - Normal Distribution Based BTV Estimates for U-232

USL 95 = 0.0653 > largest ND = 0.057

How to treat USL95 as a detected value or a nondetect?

Example 6. Consider a plutonium-241 (Pu-241) dataset collected from the three RBRAs. All but one observation are reported as NDs. The detected observation comes from surface soil sample BP-8-SUR. The detected of BP-8-SUR value = 0.437. Therefore, for Pu-241 an estimate of BTV may be considered as a ND. Based upon the available data, BTV for Pu-241 is estimated by the largest ND value = 0.349.

If one treats all NDs as detects and computes BTV estimates treating all Pu-241 activity as detects, another level of uncertainty/noise is added to the BTV estimate. It is not clear whether the BTV estimate represents a ND or a detected measurement. Treating all Pu-241 data as detects, the resulting dataset follows a normal distribution, and normal distribution based BTV estimates are shown in Figure 12 as follows. The USL95 = 0.514.

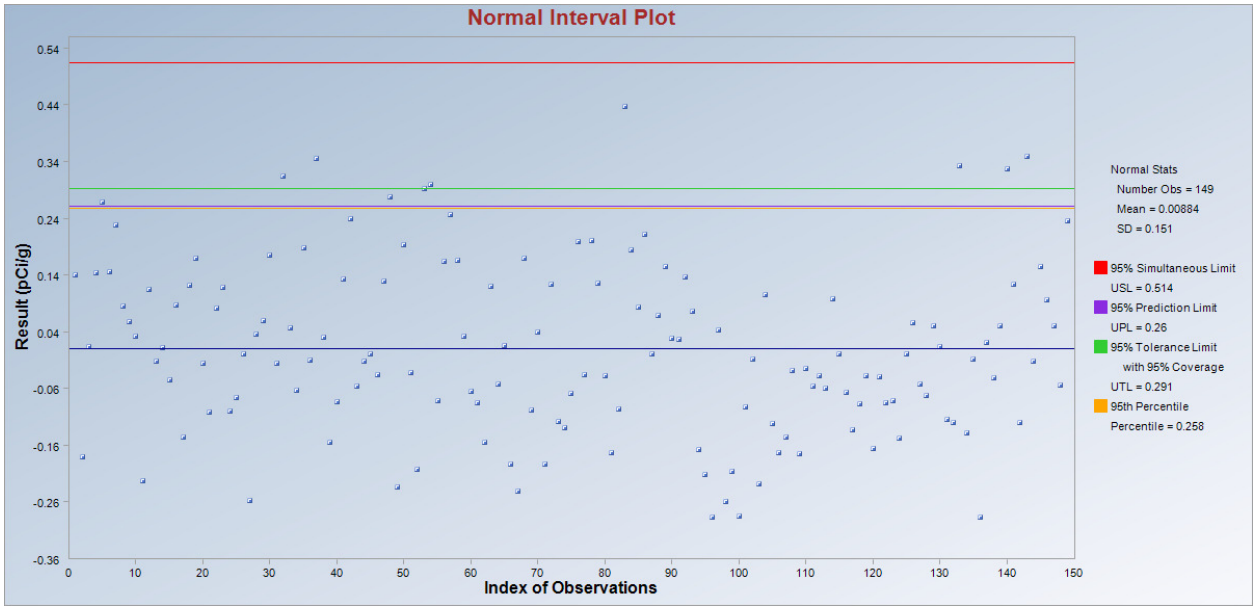

Figure 12 - Normal Distribution Based BTV Estimates for Pu-241 Activity

Example 7. Consider the americium-241 (Am-241) background dataset collected from the three RBRAs of the two formations selected for the SSFL Background Study. The dataset consists of five detects. Accounting for the ND observations, the KM method based BTV estimates are shown in Figure 13 below.

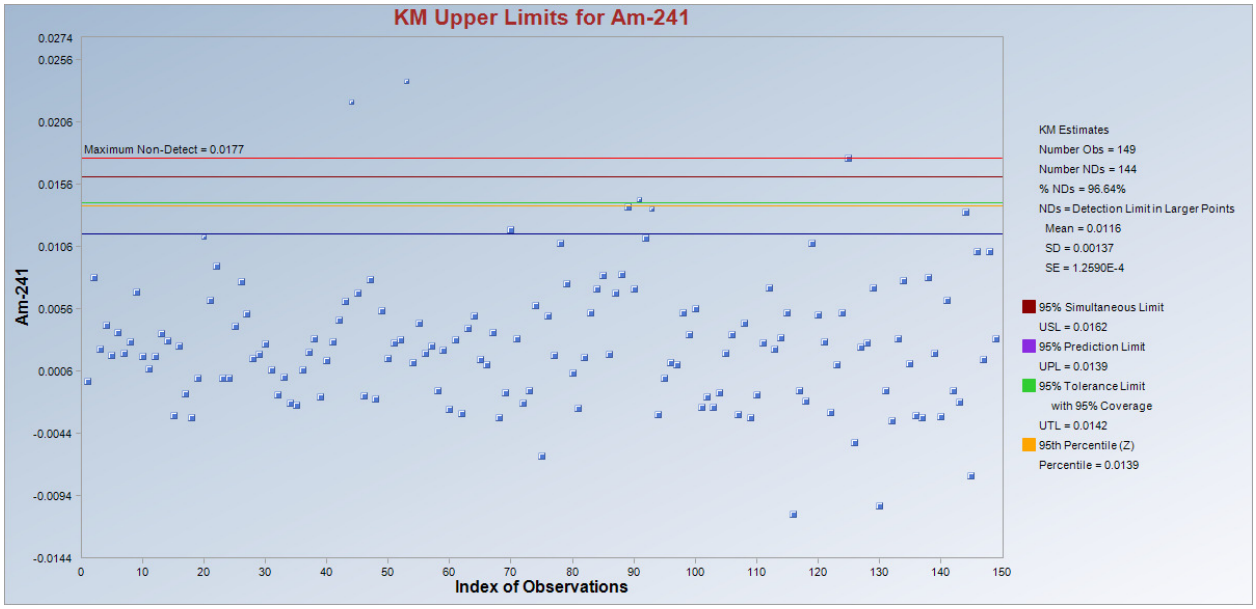

Figure 13 - KM Method Based BTV Estimates

Largest ND = 0.0177 ; and Largest Detect = 0.0239, and USL95 = 0.0162

Treating all NDs as detects, the normal distribution based and nonparametric BTV estimates are shown in Figures 14 and 15 below.

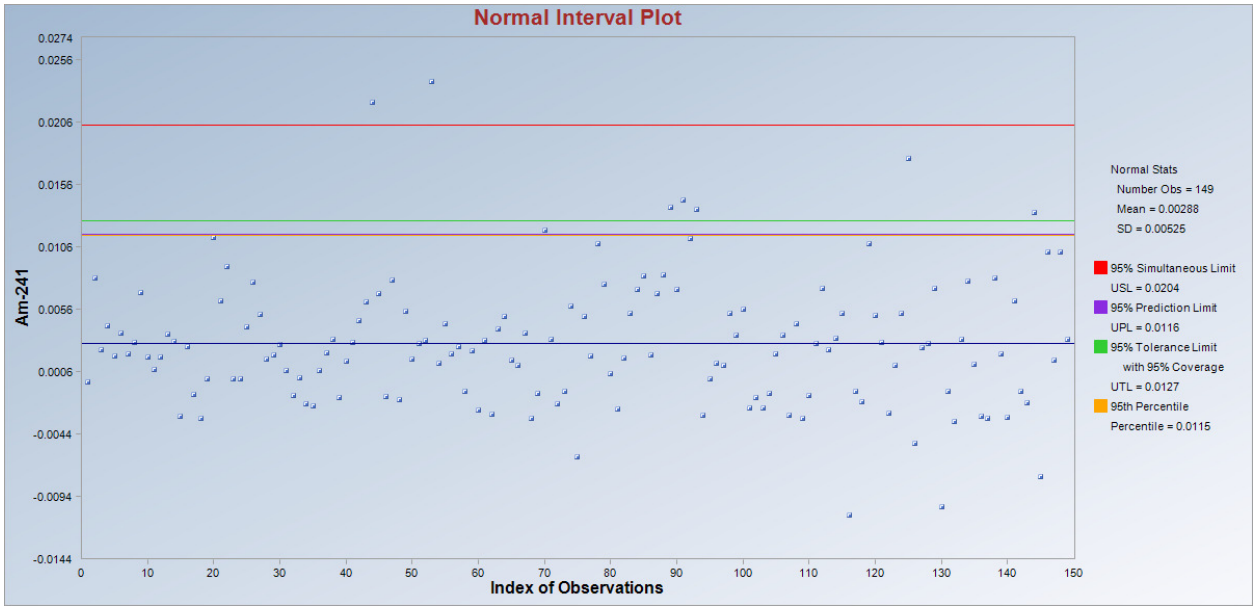

Figure 14 - Normal Distribution Based BTV Estimates (treating NDs as detects), USL=0.0204

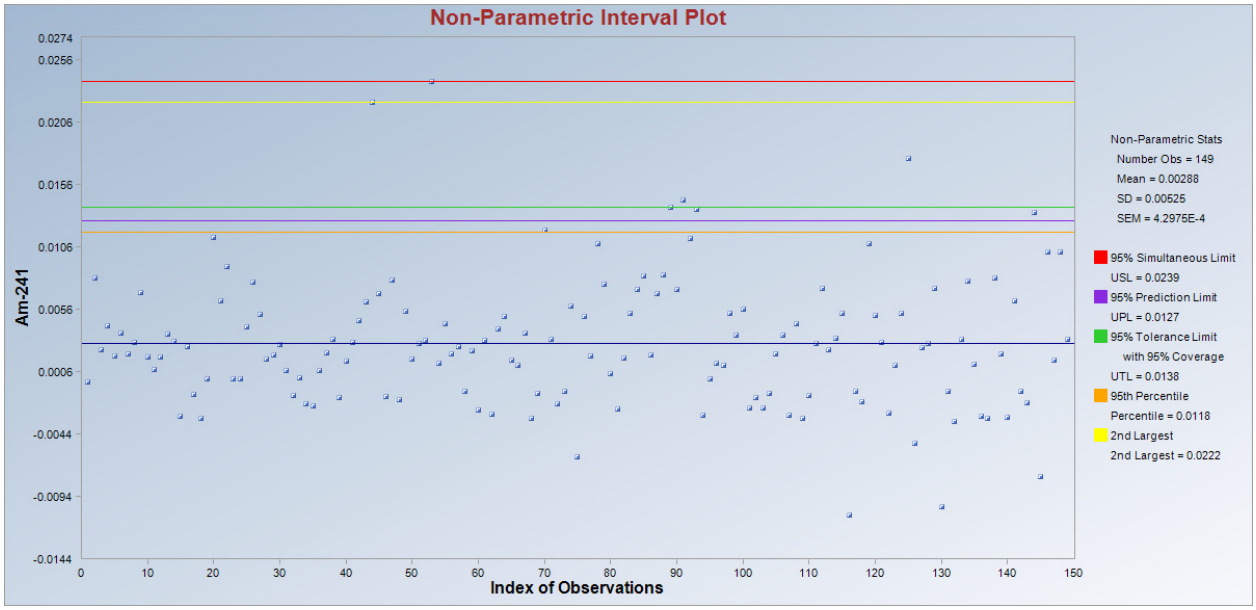

Figure 15 - Nonparametric BTV Estimates (treating NDs as detects), USL = 0.0239

5.0 REFERENCES

Barnett, V., and Lewis, T. 1994, *Outliers in Statistical Data*, 3rd ed., John Wiley and Sons, NY, 1994.

Conover W.J., 1999. *Practical Nonparametric Statistics*, Third Edition, John Wiley & Sons, NY.

David, H.A. and H.N. Nagaraja, 2003. *Order Statistics*. Third Edition. John Wiley.

Dixon, W.J. 1953, *Processing Data for Outliers*, Biometrics 9, pp.74-89.

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Gnanadesikan, R. 1997 *Methods for Statistical Data Analysis of Multivariate Observations*, 2nd ed., John Wiley and Sons, NY.

Gogolak, C.V., G.E. Powers, and A.M. Huffert, 1998. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission (NRC), NUREG-1505*. June 1998 (Rev. 1).

Hahn, J. G. and W.Q. Meeker, 1991. *Statistical Intervals. A Guide for Practitioners*. John Wiley.

Hawkins, D. M., and R.A.J. Wixley, 1986, "*A Note on the Transformation of Chi-Squared Variables to Normality*," The American Statistician, 40, 296–298.

Helsel, D.R., 2005. *Nondetects and Data Analysis*. Statistics for Censored Environmental Data. John Wiley and Sons, NY.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W., Understanding Robust and Exploratory Data Analysis, John Wiley and Sons, NY, 1983.

Johnson, R.A. and D. Wichern. 2002. *Applied Multivariate Statistical Analysis*. 6th Edition. Prentice Hall.

Kaplan, E.L. and O. Meier, 1958. *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, Vol. 53. 457-481.

Krishnamoorthy, K., T. Mathew, and S. Mukherjee, 2008. *Normal distribution based methods for a Gamma distribution: Prediction and Tolerance Interval and stress-strength reliability*.

Maronna, R.A., Martin, R.D., and Yohai, V.J. 2006, *Robust Statistics: Theory and Methods,* John Wiley and Sons, Hoboken, NJ, 2006.

Mosteller, F., and Tukey, J.W. 1977, *Data Analysis and Regression*, Addison-Wesley, Reading, MA.

Natrella, M.G., 1963. *Experimental Statistics*. National Bureau of Standards, Hand Book No. 91, U.S. Government Printing Office, Washington, DC.

Rosner, B., 1975. *On the Detection of Many Outliers.* Technometrics. Vol 17, No.2.

Rousseeuw, P.J., and Leroy, A.M. 1987, *Robust Regression and Outlier Detection*, John Wiley and Sons, NY.

Rousseeuw, P.J., and van Driessen, K. 1999, *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, Technometrics, 41, 1999, pp. 212-223.

Rousseeuw, P.J., and van Zomeren, B.C. 1990, *Unmasking Multivariate Outliers and Leverage Points*," Journal of the American Statistical Association, 85, pp. 633-651.

Scout, 2008. *Version 1.0 - A Data Analysis Program*, Technology Support Project. EPA, NERL-LV, Las Vegas, NV 89193-3478.

Singh, A., 1996. *Outliers and Robust Procedures in Some Chemometrics Applications.* The Journal of Chemometrics and Intelligent Laboratory Systems, Vol 33, 75-100.

Singh, A. and J.M. Nocerino, 1995. *Robust Procedures for the Identification of Multiple Outliers*. Handbook of Environmental Chemistry, Statistical Methods, Vol. 2.G, pp. 229-277. Springer Verlag, Germany.

Singh, A. and J.M. Nocerino, 1997. *"Robust Intervals for Some Environmental Applications"* The Journal of Chemometrics and Intelligent Laboratory Systems, Vol 37, 55-69.

Singh, A. and J.M. Nocerino, 2002. *Robust Estimation of the Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations,* Vol. 60, pp. 69-86.

Singh A., A.K. Singh, and M. Engelhardt, 1997. *The Lognormal Distribution in Environmental Applications.* Office of Research and Development. EPA/600/S-97/006. http://www.epa.gov/esd/tsc/images/256cmb04lognorm.pdf

Singh, A., A.K. Singh, and R. Iaci, 2002. *Computation of the Exposure Point Concentration Term Using a Gamma Distribution.* EPA/600/R-02/084. October, 2002. http://www.epa.gov/esd/tsc/images/289cmb02.pdf

Singh, A., R. Maichle, and S. Lee, 2006. *On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations.* EPA/600/R-06/022, March, 2006. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

Singh, A., Maichle, R., Armbya, N., and Singh, A.K. 2009. *Scout 2008, Version 1.00.01, User Guide.* Parts I, II, III, and IV. EPA/600/R-08/038. February, 2009. http://www.epa.gov/esd/databases/scout/abstract.htm

Singh, Anita, Maichle, R., and Armbya, N. 2009. Software - *Scout 2008, Version 1.00.01. A Statistical Software package for the classical and Robust Analysis of Univariate and Multivariate Data Sets with and without Non-detect Observations.* Download software from: http://www.epa.gov/esd/databases/scout/abstract.htmSingh, Anita, Maichle, R., and Armbya, N. 2010. Software - Scout 1.1, Version 1.0. *A Statistical Software package for the classical and Robust Analysis of Univariate and Multivariate Data Sets with and without Non-detect Observations.* Submitted to NERL-EPA, Las Vegas in September 2010.

Tukey, J.W. *Exploratory Data Analysis.* Addison-Wesley Publishing Company, Reading, MA, 1977.

U.S. Environmental Protection Agency (EPA), 1989. *Statistical Analysis of Ground-water Monitoring Data at RCRA Facilities*. Interim Final Guidance. Washington, DC: Office of Solid Waste. April, 1989.

U.S. Environmental Protection Agency (EPA), 1992. *Statistical Analysis of Ground-water Monitoring Data at RCRA Facilities*. Addendum to Interim Final Guidance. Washington DC: Office of Solid Waste. July, 1992.

U.S. Environmental Protection Agency (EPA), 2000. *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)*. NUREG-1575, Revision 1, EPA 402-R-97-016, Rev.1, DOE/EH-0624, Rev. 1. Office of Research and Development, Quality Assurance Division, Washington DC. Available at http://www.epa.gov/radiation/marssim/ or from http://bookstore.gpo.gov/index.html (GPO Stock Number for Revision 1 is 052-020-00814-1).

U.S. Environmental Protection Agency (EPA), 2002. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA 540-R-01-003-OSWER 9285.7-41. September, 2002.

U.S. Environmental Protection Agency (EPA), 2006. *Data Quality Assessment: Statistical Methods for Practitioners*, EPA QA/G-9S. EPA/240/B-06/003. Office of Environmental Information, Washington, DC. Download from: http://www.epa.gov/quality/qs-docs/g9s-final.pdf

U.S. Environmental Protection Agency (EPA), 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance*. EPA 530-R-09-007. 2009.

U.S. Environmental Protection Agency (EPA), 2010a. ProUCL Version 4.00.05 User Guide (Draft). EPA/600/R-07/041. May, 2010. http://www.epa.gov/osp/hstl/tsc/software.htm

U.S. Environmental Protection Agency (EPA), 2010a. ProUCL Version 4.00.05 Technical Guide (Draft). EPA/600/R-07/041. May, 2010. http://www.epa.gov/osp/hstl/tsc/software.htm

U.S. Environmental Protection Agency (EPA), 2011. ProUCL 4.1. *A Statistical Software for Environmental Applications for Data Sets with and without nondetect observations*. National Exposure Research Lab, EPA, Las Vegas Nevada. March, 2011. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

U.S. Environmental Protection Agency (2011). ProUCL 4.1. *A Statistical Software for Environmental Applications for Data Sets with and without nondetect observations*. National Exposure Research Lab, EPA, Las Vegas Nevada, March 2011. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

Wilson, E. B., and M.M. Hilferty, 1931. "*The Distribution of Chi-Squares*," Proceedings of the National Academy of Sciences, 17, 684–688.

Wilks, S.S (1963). *Multivariate Statistical Outliers*. Sankhya, 25, 407-226.